

Appl. No. 09/940,188
Amdt. dated April 6, 2004
Reply to Office action of Oct. 6, 2003

REMARKS/ARGUMENTS

Claims 1 - 23 remain in this application. Claim 24 has been canceled. Claims 1, 7 and 12 have been amended.

Objections to Claims 12 and 24

Claim 24 was objected to as being substantial duplicate of claim 11. In response, claim 24 has been canceled.

Claim 12 was objected to because of a typing/grammar error. Correction has been made. Claim 7 has been similarly changed.

Rejection of Claims 1-24 Under 35 U.S.C. §102(e)

Claims 1-24 were rejected under 35 U.S.C. §102(e) as being anticipated by U.S. Patent No. 6,463,430 (Brady et al.). Applicant respectfully traverses the rejection.

Claim 1 has been amended to add language describing a search metric and crawl metric, based on the applicant's specification. See the specification, page 12, lines 10-14. This amendment is not believed to change the meaning of the claim, since it is based on the language in the specification which explains the terminology used in the claim. Accordingly, the amendment has been made only for the purpose of clarification and not limitation.

Claim 1, as amended, includes:

1. (previously amended) A method of focused crawling, comprising:
accessing a query input, the query input including at least a first query part and a second query part;
crawling a plurality of documents, at least some of the plurality of documents including links to each other, the crawling at least partly guided by a crawl metric, the crawl metric at least partly determined by a mechanism and by the first query part; and

Appl. No. 09/940,188
Amdt. dated April 6, 2004
Reply to Office action of Oct. 6, 2003

returning target documents, the target documents being relevant to the second query part, the target documents found from the plurality of crawled documents, the target documents returned at least partly based on a search metric, the search metric at least partly determined by the mechanism and by the second query part part;
wherein the crawl metric comprises a metric that quantifies priority for crawling links emanating from a certain document within context of the focused crawling and
wherein the search metric comprises a metric that quantifies relevance or importance of a document to the query input.

It is believed that Brady fails to anticipate claim 1. For example, Brady fails to teach crawling at least partially guided by a crawl metric, and documents being returned at least partially based on a search metric.

For a crawl metric, the Office Action points to the definition of Retrieval Priority List in Brady. However, looking at the definition of Retrieval Priority List in Brady, it can be seen that this is a different concept from the claimed invention. Brady states:

As used herein, the term "Retrieval Priority List" refers to repository of hypertext links, URL addresses, or the like, used in retrieving documents from a network of documents, such as, the Internet, World Wide Web, LANs, intranets, or the like.

Brady, col. 4, li. 66 - col. 5, li. 3. In contrast, the claimed crawl metric comprises a metric that quantifies priority for crawling links emanating from a certain document within context of the focused crawling. Thus, Brady does not teach the claimed crawl metric.

From claim 1, it can also be seen that the crawl metric is a separate concept from the search metric. The crawl metric comprises "a metric that quantifies priority for crawling links emanating from a certain document within context of the focused crawling," while the search metric comprises "a metric that quantifies relevance or importance of a document to the query input." See claim 1, as amended. Thus, claim 1 teaches crawling at least partially guided by the crawl metric, and returning target documents at least partly based on the search metric.

Appl. No. 09/940,188
Amdt. dated April 6, 2004
Reply to Office action of Oct. 6, 2003

In contrast, for returning target documents based on a search metric, the Office Action cites Fig. 11 of Brady and the same concept discussed above of the Retrieval Priority List. See Office Action at 4. Brady does not teach the separate use of crawl metric and search metric as claimed in claim 1. Accordingly, Brady does not teach crawling at least partially guided by the crawl metric, and returning target documents at least partly based on the search metric, as these concepts are understood in claim 1. In particular, Brady does not teach such using the separate concepts crawl metric and search metric where the crawl metric comprises a metric that quantifies priority for crawling links emanating from a certain document within context of the focused crawling, and the search metric comprises "a metric that quantifies relevance or importance of a document to the query input."

Further, the query input claimed in claim 1 potentially includes input for both crawl metric and search metric, which are separate as discussed above. See claim 1 indicating that the crawl metric is at least partly determined by a mechanism and by the first query part, and that the search metric is at least partly determined by the mechanism and by the second query part. The Office Action has not shown any teaching in Brady of using such parts of a query for a separate crawl metric and search metric. Rather, the Office Action indicates that a supposed crawl metric of Brady is determined by a "first mechanism," (emphasis added) and that a supposed search mechanism is determined by a "second mechanism," (emphasis added) without mentioning use of a first query part or second query part. Thus, for this additional reason, Brady fails to anticipate claim 1.

Thus, it is believed that Brady fails to anticipate claim 1. Removal of the rejection is therefore respectfully requested.

Appl. No. 09/940,188
Amdt. dated April 6, 2004
Reply to Office action of Oct. 6, 2003

Claims 3, 5 and 10 were rejected based on similar reasoning as to claim 1. It is therefore believed that the rejection of such claims should also be removed.

Other pending claims depend directly or indirectly from claims 1, 3, 5 or 10. Since such parent claims are believed patentable, these dependent claims are also believed patentable. Removal of the rejection of such claims is therefore respectfully requested.

Thus, for the reasons discussed above, it is believed that all claims in the application are allowable and passage of the claims to allowance is respectfully requested. For completeness, however, the following is discussion of some other aspects of the rejection in the Office Action, where it is believed that Brady has not been applied correctly to the claims of the present application.

For example, the Office Action, at 4-5, items e) - h) refers to items related to the "taxonomy" of Brady, equating this with a graph. It is believed that such portions of Brady do not teach the evaluation of the linked structure as taught in the claimed invention, for example, as taught in claim 10. For example, claim 10 refers to "generating a graph of the first plurality of documents." In item e), page 4, the Office Action refers to generating a graph of the first plurality of documents and points to taxonomy in Brady item 1301, Fig. 13 and column 24, li. 5-19 and col. 23, li. 32-40. It is believed that this application of Brady is incorrect. It is noted that Brady indicates that "subject taxonomy" refers to a subject area for which information is gathered and categorized. See Brady, col. 4, li. 57-59. A taxonomy is a classification, not a graph. Brady uses the term taxonomy in the sense of determining relevancy of documents to the taxonomy. See Brady, Abstract. It would not make sense at the same time for the taxonomy to

Appl. No. 09/940,188
Amdt. dated April 6, 2004
Reply to Office action of Oct. 6, 2003

be a graph of the documents and for the documents to have a relevancy to the taxonomy. Thus, Brady is not teaching generating a graph of the documents.

The portion of column 24 cited in the Office Action regarding taxonomy refers to managing a database over the Internet. There is no teaching of the generation of a graph of the documents. The portion of column 23 cited in the Office Action for taxonomy refers to developing taxonomies for organizing retrieved resources, and placing retrieved resources into specific categories of the taxonomy. Thus, the taxonomy of Brady provides a way to organize retrieved resources, and the taxonomy has categories, but with such language Brady does not teach generating a graph of the documents.

As noted, Brady does not teach the claimed generation of a graph. Thus, there cannot be teaching of propagating weights through such a graph, since there is no generation of such a graph. The Office Action points to steps 804-812, Fig. 8 of Brady. However, such section provides no teaching of propagation of weights through a graph of a plurality of documents. Rather, here Brady teaches classification of a web page. The classifier selects a taxonomy category for a web page, but such selection and the associated scoring of the respective page is not propagating weights through a graph of a plurality of documents.

Thus, for these additional reasons, it is believed that the rejection under 35 U.S.C. §102(e) is incorrect.

Additionally, it is not believed that the Office Action is correct in the statement that "relevance includes importance is defaulted by the page ranking technique." Relevance does not necessarily refer to ranking.

Appl. No. 09/940,188
Amdt. dated April 6, 2004
Reply to Office action of Oct. 6, 2003

As to claims 8 and 13, it is not clear how the cited portion of Brady would apply to the claims. Such claims refer to hierarchical, heading and content levels. The cited portion of Brady, col. 22, lines 44-46 refer to searches that identify resources, such as web documents, web pages and the like, that match the user query. There appears to be no teaching that pertains to these claims here.

It is believed that the other discussion of Brady in the Office Action is not applicable to the pending claims for reasons similar to those discussed above.

Thus, it is believed that claims 1-23 are allowable and removal of the rejection is respectfully requested.

Appl. No. 09/940,188
Amdt. dated April 6, 2004
Reply to Office action of Oct. 6, 2003

CONCLUSION

Applicants submit that the instant application is in condition for allowance. Should the Examiner have any questions, the Examiner is requested to contact the undersigned attorney.

The Commissioner is authorized to charge any additional fees which may be required, including petition fees and extension of time fees, to Deposit Account No. 23-2415 (Docket No. 25961.706).

Respectfully submitted,

WILSON SONSINI GOODRICH & ROSATI

Date: April 6, 2004


George A. Wilkman, Reg. No. 41,378

650 Page Mill Road
Palo Alto, CA 94304
(650) 595-3995
Customer No. 021971